The Multiple Hypothesis Tracker Derived from Finite Set Statistics

Edmund Førland Brekke
Dept. of Engineering Cybernetics
NTNU
7034 Trondheim, Norway
Email: edmund.brekke@ntnu.no

Mandar Chitre
ARL, Tropical Marine Science Institute
National University of Singapore
14 Kent Ridge Road, Singapore 119222
Email: mandar@arl.nus.edu.sg

Abstract—**The multiple hypothesis tracker (MHT) has historically been considered a gold standard for multi-target tracking. In this paper we show that the key formula for hypothesis probabilities in Reid's MHT can be derived from the modern theory of finite set statistics (FISST) insofar as appropriate assumptions (Poisson models for clutter and undetected targets, no target-death, linear-Gaussian Markov target kinematics) are adhered to.**

I. INTRODUCTION

In target tracking the aim is to establish tracks on all targets in a surveillance region, and maintain these tracks for as long as the corresponding targets exist or remain inside the surveillance region. The standard multitarget tracking problem can be decomposed into the two subproblems of *filtering* and *data association*. Filtering concerns the estimation of a target's kinematic state from a time series of measurements. Data association concerns how one determines which measurements originate from which targets, and which measurements are to be considered useless *clutter*. Many of the tasks carried out in a surveillance system are most appropriately understood in the context of data association. For example, establishing tracks on new targets is fundamentally a data association problem, since a decision to establish a new track amounts to a decision regarding the origin of a sequence of measurements.

In early developments, data association was viewed as a problem of determining an optimal association hypothesis [1]. Reid's multiple hypothesis tracker (MHT) took this one step further by recognizing that every possible association hypothesis could be assigned a posterior probability [2], which could be propagated recursively in time according to (16) in [2]. Reid's MHT has historically been regarded as the optimal approach to data association, although the exact nature of this presumed optimality was never stated clearly.

Finite set statistics (FISST) was introduced as a more systematic alternative to multi-target tracking than the MHT paradigm [3]. In FISST, both targets and measurements are generally treated as random finite sets, i.e., as set-valued random variables. This allows one to express a Bayes-optimal solution to the full multi-target tracking problem, including data association, using a single prediction equation and a single update equation. This is known as the multi-target Bayes filter. FISST came with several approximations of the multi-target Bayes filter, such as the probablity hypothesis density (PHD) filter [4] and the cardinalized PHD (CPHD) filter [5].

The brute force multi-target Bayes filter itself received little attention to begin with, as it clearly was intractable. In 2015, however, Jason Williams [6] showed that its output can be factorized into a Poisson component and multi-Bernoulli mixture, which can be viewed as a sum over association hypotheses which are somewhat similar to those used in Reid's MHT. The focus in [6] was on deriving approximations of the multi-target Bayes filter similar to the joint integrated probabilistic data association (JIPDA) [7], while the relationship to MHT methods only was treated in cursorial terms.

The main contribution of this paper is to show that key ingredients in Reid's MHT can be derived from the representation of the multi-target Bayes filter that was proposed in [6]. In particular, we show that Reid's formula follows from the expressions in [6].

This analysis has important ramifications. It is often believed that the two paradigms of MHT and FISST are mutually irreconcilable. If so, one could suspect that fundamental flaws should hide in at least one of the two paradigms, and such a suspicion may cause reluctance towards exchange of ideas between the two schools. On the other hand, if Reid's MHT can be shown to fit nicely within the theoretical framework of FISST, then such concerns are much less warranted. The relationship between FISST and MHT has been discussed in many papers, including [8], [9] and [10]. However, the relationship between FISST and Reid's formula has not yet been fully elucidated.

This paper is inspired by the authors' previous work on MHT-techniques for simultaneous localization and mapping (SLAM) [11, 12]. It borrows much notation from [12], which again was heavily influenced by preprints of [6] and the PhD thesis [13].

The paper is organized as follows. We introduce important notations in Section II. The two kinds of association hypotheses that we use are discussed in Section III. A brief introduction to FISST is given in Section IV. Modeling assumptions are stated in Section V. The main developments of the paper find place in Section VI, while the special case of Gaussian-linear kinematics follows in Section VII, before the conclusion in Section VIII.

II. NOTATION

In this section we provide an overview of notation used, before we properly introduce association hypotheses, FISST and assumptions employed in the subsequent sections.

A. Probability density functions and Poisson intensities

Various pdf's that we encounter include the target Markov pdf $f_x(x_k|x_{k-1})$, the single-target likelihood $f_z(z_k \,|\, x_k)$, the single-track predicted pdf $f_{k|k-1}^{i,a_k^i}(x_k)$ and the single-track posterior pdf $f_k^{i,a_k^i}(x_k)$. Poisson-processes are given in terms of their intensities, such as $\lambda(z)$, $v(x)$ or $\mu(x)$.

B. Probability-generating functionals

While random finite sets (RFS's) can be described in term of multiobject densities, we will mainly use the transform-domain representation in terms of probability generating functionals (p.g.fl.'s) [3] in this paper. The general notation for p.g.fl.'s are using capital G or H and square brackets, e.g., $G_{k-1}[h]$. We make extensive use of linear functionals which correspond to pdf's or Poisson intensities. These are always written using lowercase Latin or Greek letters. Test functions of functionals are always denoted h. Variables bound within the scope of a functional are marked by a tilde. Such variables, e.g., $\tilde{x}_k$, are always integrated out as part of the functional. For example,

$$f_k^{i,a^i}[h] = \int h(x) f_k^{i,a^i}(x)\mathrm{d}x$$

$$v[f_z(z \,|\, \tilde{x})] = \int f_z(z \,|\, x)v(x)\mathrm{d}x, \;\; \text{etc.}$$

C. Time-, track- and measurement-indices

Time is accounted for by the time index k. The list of time steps $(l, l+1, \ldots, k-1, k)$ is in shorthand notation written as $l : k$. Association hypotheses are denoted a and b. The sets of such hypotheses at time step k are denoted $\mathcal{A}_k$ and $\mathcal{B}_k$, respectively. Measurement indices are denoted j, while track indices are denoted i. The notation a_k^i indicates the measurement claimed by track number i at time step k according to a. A notation such as $a_{1:k-1}^{1:n_{k-1}}$ indicates the matrix of track-to-measurement associations for tracks 1 to n_{k-1} for time steps 1 to $k-1$ according to a. Notation for association hypotheses is further elaborated in the next section.

III. ASSOCIATION HYPOTHESES

Any attempt at expressing the full multi-object density for conventional MTT problems must necessarily invoke association hypotheses in one way or another. However, there is limited consensus on how the fundamental concepts of *association hypotheses* and *tracks* should be understood and defined. In this paper, a track is defined as a temporal sequence of measurement indices,[1] possibly including the zeroth measurement, which represents misdetection or non-existing target, depending on context.

[1]This is called a single-target hypothesis in [6].

A. Williams' hypotheses

The hypothesis-representation proposed in [6], hereafter referred to as W-hypotheses, was described as similar to Kurien's track-oriented MHT [14], since it utilizes similar track trees. It differs from conventional MHT representations, including Kurien's, since it in some cases allows the possibilities of both existence and non-existence of the target to be maintained within a single node in this track tree. In particular, when a new track tree is established, the possibilities that the root measurement is a measurement of the hypothesized target, and that the root measurement is just clutter, are both accounted for by the same root node.

For notational convenience, and in order to explore similarities with Reid's MHT, we use a matrix-based notation for this representation, originally proposed in [12]. The matrices are of dimension $k \times n_k$, where k is the last time step and n_k is the number of measurements so far processed. The bookkeeping is done by means of the linear index function $L(j; k) = j + \sum_{l=1}^{k-1} m_l$ with inverse $L^{-1}(i; k) = i - \sum_{l=1}^{k-1} m_l$. What L does is to give us the index of the track with origin in the measurement z_k^j, while its inverse gives the identity of the original measurement in track number i. The number of tracks at time k is $n_k = L(m_k; k)$. We can then define the set $\mathcal{A}_k$ of W-hypotheses at time k as the set of mappings[2] $a : \{1, \ldots, n_k\} \to \{0, \ldots, m_1\} \times \ldots \times \{0, \ldots, m_k\}$ such that

R1: For any measurement index $j \in \{1, \ldots, m_k\}$ there exists one track index $i \in \{1, \ldots, n_k\}$ such that $a_k^i = j$.

R2: For any measurement index $j \in \{1, \ldots, m_k\}$ there exists at most one track index $i \in \{1, \ldots, n_k\}$ such that $a_k^i = j$.

R3: There exists a hypothesis $\tilde{a}$ in the previous hypothesis set $\mathcal{A}_{k-1}$ such that $a_{1:k-1}^{1:n_{k-1}} = \tilde{a}$.

R4: For any track index $i \in \{1, \ldots, n_k\}$, if there exists no $l < k$ such that $a_l^i > 0$, and $i \leq n_{k-1}$, then $a_k^i = 0$.

R5: For any track index $i \in \{1, \ldots, n_k\}$, if there exists no $l < k$ such that $a_l^i > 0$, and $i > n_{k-1}$, then $a_k^i = L^{-1}(i; k)$.

Requirement R1 means that all measurements must be accounted for. Requirement R2 means that tracks cannot share measurements. Requirement R3 means that any current hypothesis must have a parent hypothesis, i.e., that the collection of hypotheses is expanded in a recursive manner. Requirement R4 means that if the track with origin in z_k^i fails to claim z_k^i, then this track represents a non-existing target for all future. Requirement R5 means that any new track with index i at time step k must originate with the corresponding measurement at time step k.

B. Reid's hypotheses

Any set of W-hypotheses gives rise to a larger set of hypotheses $b \in \mathcal{B}_k \supseteq \mathcal{A}_k$ of the kind that was used by Reid in [2]. We refer to these hypotheses as R-hypotheses. For any W-hypotheses a, the corresponding set of R-hypotheses is obtained by relaxing the requirement R1. In other words,

[2]Such a mapping between integers is clearly isomorphic to a corresponding $k \times n_k$-matrix.

an R-hypothesis does not necessarily need to account for all measurements.

REMARK 1. Reid's seminal paper [2] developed the concept of association hypotheses through 3 successive refinements of the outcome space, referred to as Number, Configuration and Assignment. The *Number* event concerns how many measurements are associated with existing targets, newborn targets and with clutter. The *Configuration* event concerns how the set of measurements is partitioned into subsets associated with each of these three sources. The *Assignment* event concerns "the specific source of each measurement which has been assigned to be from some previously known target". These Assignment events are identical to the R-hypotheses as defined above.

C. Illustration of the hypothesis formalisms

Let us take a closer look at how W-hypotheses and R-hypotheses arise in the scenario illustrated in Figure 1. At $k = 1$ we have a single W-hypothesis:

$$\mathcal{A}_1 = \left\{ \begin{bmatrix} 1 \end{bmatrix} \right\}.$$

At $k = 2$ we get 3 W-hypotheses:

$$\mathcal{A}_2 = \left\{ \left[\begin{array}{c|cc} 1 & 0 & 0 \\ 0 & 1 & 2 \end{array} \right], \left[\begin{array}{c|cc} 1 & 0 & 0 \\ 1 & 0 & 2 \end{array} \right], \left[\begin{array}{c|cc} 1 & 0 & 0 \\ 2 & 1 & 0 \end{array} \right] \right\}.$$

These can be expanded into 8 R-hypotheses:

$$\mathcal{B}_2 = \left\{ \left[\begin{array}{c|cc} 1 & 0 & 0 \\ 0 & 0 & 0 \end{array} \right], \left[\begin{array}{c|cc} 1 & 0 & 0 \\ 0 & 1 & 0 \end{array} \right], \left[\begin{array}{c|cc} 1 & 0 & 0 \\ 0 & 0 & 2 \end{array} \right], \right.$$
$$\left[\begin{array}{c|cc} 1 & 0 & 0 \\ 0 & 1 & 2 \end{array} \right], \left[\begin{array}{c|cc} 1 & 0 & 0 \\ 1 & 0 & 0 \end{array} \right], \left[\begin{array}{c|cc} 1 & 0 & 0 \\ 1 & 0 & 2 \end{array} \right],$$
$$\left. \left[\begin{array}{c|cc} 1 & 0 & 0 \\ 2 & 0 & 0 \end{array} \right], \left[\begin{array}{c|cc} 1 & 0 & 0 \\ 2 & 1 & 0 \end{array} \right] \right\}.$$

It is evident that the larger amount of hypotheses in Reid's MHT is due to its method for handling new tracks. In practice, this problem can be mitigated by using a track-oriented MHT [14], which only enumerates all these tracks, and not all the generated hypotheses, and/or by using a track initialization method that works outside of the MHT. It is also evident that for every R-hypothesis there is a unique underlying W-hypothesis. We prove this later in the paper.

D. Discussion of other representations

Notable alternative approaches to the concept of association hypotheses include [15] and [16].

In [15], a distinction was made between *data-to-data hypotheses* and *target-to-data hypotheses*. Data-to-data hypotheses concern how measurements at one time step are associated to measurements at other time steps. Target-to-data hypotheses concern how measurements are generated by specific targets. Reid's Assignment events are the same as data-to-data hypotheses, while target-to-data hypotheses provide a fourth refinement of the outcome space. Target-to-data hypotheses are frequently encountered in [3], since many multiobject densities are conveniently described in terms of such entities. We argued in [11] that association hypotheses with well-defined posterior probabilities can be defined as equivalence
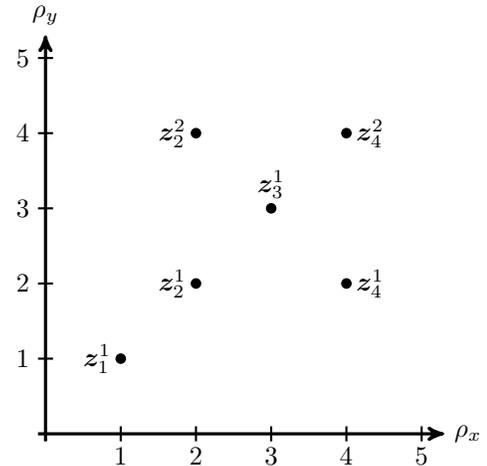

Fig. 1: Illustratory scenario.

classes of such entities, under the equivalence relation of permutation-equivalence.

In [16], the multi-target Bayes filter was generalized and formulated in terms of *labeled* random finite sets. The resulting association hypotheses are similar to W-hypotheses, but include explicit *target* labels in addition to measurement histories. The factorization of [6] is not present in [16], implying that the formalism in [16] involves a significantly larger amount of association hypotheses than in [6]. For standard tracking applications, the utility of explicit labels is not obvious: While neither [2], [6], nor [15] include explicit labels on targets, these approaches are nevertheless also able to provide labels on *tracks* in an implicit manner. We will come back to this topic towards the end of Section VI-D.

IV. KEY CONCEPTS IN FISST

FISST is a Bayesian formulation of point process theory. An RFS is essentially the same as a simple point process [4]. FISST provides rules for how the random properties of RFS's can be represented in terms of multiobject densities. This allows us to extend Bayesian probability calculus from the conventional domain of random vectors to the more complex domain of RFS's.

Multiobject densities are functional derivatives (see the next paragraph) of belief-mass functions, or more generally of p.g.fl.'s. The belief-mass function of a random finite set Ξ is the probability $\beta_\Xi(S) = Pr(\Xi \subseteq S)$. Here S is a subset of the base space, i.e., if $x \in \Xi$ and $x \in \mathbb{R}^d$, then $S \subseteq \mathbb{R}^d$. FISST utilizes the so-called set integral [3 p. 361]. Let $f(X)$ be a set function, i.e., a function of the realization X of the finite set Ξ. Then the set integral of $f(X)$ over S is defined as

$$\int_S f(X)\delta X = \sum_{n=0}^{\infty} \frac{1}{n!} \int_{S \times \ldots \times S} f(\{x^1, \ldots, x^n\})\mathrm{d}x^1 \ldots \mathrm{d}x^n.$$

Here n is the cardinality of X, and $f(\{x^1, \ldots, x^n\}) = f(X)$ under the constraint that $|X| = n$. A multiobject density

$f_\Xi(X)$ is a function which produces a non-negative real number from any realization X of Ξ, and which normalizes to one under the set integral: $\int f_\Xi(X)\delta X = 1$. The multiobject density $f_\Xi(X)$ is related to the belief-mass function according to the requirement that $\beta_\Xi(S) = \int_S f_\Xi(X)\delta X$ for all $S \subseteq \mathbb{R}^d$.

Belief-mass functions are special cases of p.g.fl.'s. The p.g.fl. of a random set Ξ is defined as

$$G_\Xi[h] = \int h^X f_\Xi(X)\delta X \tag{1}$$

where $h : \mathbb{R}^d \to [0,\infty)$ is a test-function, and the notation h^X signifies the product $\prod_{x\in X} h(x)$. According to the fundamental theorem of multiobject calculus [3 p. 384] we can recover the multiobject density from the p.g.fl. by

$$f_\Xi(X) = \frac{\delta G_\Xi}{\delta X}[0]. \tag{2}$$

Here $\frac{\delta G_\Xi}{\delta X} = \frac{\delta^n G_\Xi}{\delta x^1 \cdots \delta x^n}$ is the iterated functional derivative of G_Ξ in the directions of the delta-functions $\delta_{x^1}(\cdot), \ldots, \delta_{x^n}(\cdot)$ under the condition that $X = \{x^1, \ldots, x^n\}$. These derivatives are found as the limit values

$$\frac{\delta G}{\delta x}[h] = \lim_{\epsilon \searrow 0} \frac{G[h + \epsilon\delta_x] - G[h]}{\epsilon}. \tag{3}$$

Thus, p.g.fl.'s provide a transform-domain for the manipulation of multiobject densities. Many properties of p.g.fl.'s are listed in [3], and these properties often make manipulation of p.g.fl.'s easier than manipulation of the corresponding multiobject densities. For example, the p.g.fl. of the union of two independent random sets is simply the product the respective p.g.fl.'s of those sets. The corresponding relationship for multiobject densities is more complicated.

A. Important random sets and their p.g.fl.'s

The Poisson RFS with intensity $\lambda(x)$ has multiobject density and p.g.fl. given by

$$f(X) = \exp(-\lambda[1]) \prod_{x\in X} \lambda(x) \tag{4}$$

$$G[h] = \exp(\lambda[h-1]). \tag{5}$$

The p.g.fl. of the singleton set is referred to as a linear functional. If $X = \{x\}$ with probability one, and x has the pdf $f(x)$, then the p.g.fl. of X is

$$f[h] = \int f(x)h(x)\mathrm{d}x. \tag{6}$$

A Bernoulli RFS with existence probability r and existence-conditioned pdf $f(x)$ has multiobject density and p.g.fl.

$$f(X) = \begin{cases} 1-r & \text{if } X = \emptyset \\ rf(x) & \text{if } X = \{x\} \end{cases} \tag{7}$$

$$G[h] = 1 - r + rf[h]. \tag{8}$$

Notice that the Bernoulli p.g.fl. reduces to a linear p.g.fl. in the limit case of unity existence probability.

A union of independent Bernoulli random sets is known as a multi-Bernoulli RFS. Its p.g.fl. is simply the product of its constituting Bernoulli p.g.fl.'s, while its multiobject density

has a more complicated expression (see (11) in [6] or page 368 in [3]).

An RFS whose p.g.fl. is a weighted sum of multi-Bernoulli p.g.fl.'s is known as a multi-Bernoulli mixture. If the sum goes over W-hypotheses we refer to this as a W-MBM. If the sum goes over R-hypotheses we refer to this as an R-MBM.

V. ASSUMPTIONS

During estimation cycle number k we assume that all ξ_{k-1} targets that existed at time k with states x^i_{k-1}, $i = 1, \ldots, \xi_{k-1}$ remain from the previous cycle. At time k each of these targets have densities

$$f(x^i_k) = \int f_x(x^i_k | x^i_{k-1}) f(x^i_{k-1}) \mathrm{d}x^i_{k-1}$$

where $f_x(x^i_k | x^i_{k-1})$ is the kinematic transition pdf and $f(x^i_{k-1})$ is the posterior pdf at time $k-1$. In addition to these targets, new targets may appear at time k according to a Poisson process with intensity $\mu(x)$. We will discuss the relationship between this bona-fide birth model and Reid's more vague assumptions on this point in more detail in Section VII-B. Once a target exists, it cannot die.

At time k we receive m_k measurements $z^1_k, \ldots, z^{m_k}_k$. Given that measurement z^j_k is generated by target x^i_k, its kinematic likelihood is given by $f_z(z^j_k | x^i_k)$. Any target generates a single measurement with constant probability P_D, otherwise it is unobserved. An unknown number φ_k of the m_k measurements are clutter measurements generated by a Poisson process whose intensity λ is constant and non-zero everywhere in the sensor's field of view.

Furthermore, standard independence assumptions apply (see. e.g. [2], [3] or [17]).

VI. THE MULTI-TARGET BAYES FILTER WRITTEN IN TERMS OF REID'S HYPOTHESES

This section contains the main contribution of the paper, which is to show that we can recover Reid's recursive hypothesis probability formula [2] from Williams' formulation of the multi-target Bayes filter [6]. Before delving into the technical details, we briefly sketch the main steps of this argument:

A Any W-MBM can be expanded into an R-MBM. Thus we are free to formulate the multi-hypothesis nature of the original prior as an R-MBM.

B We can obtain the predicted R-MBM using the same machinery as for the W-MBM.

C The measurement update transforms the R-MBM into a W-MBM.

D The only components with non-integer existence probability in this W-MBM are tracks on new targets. Following the premise of Step A we can expand the products over new targets to get a new R-MBM. We can then find the probability of each new R-hypothesis as some factor times the probability of its parent R-hypothesis.

In the remainder of this section we elaborate and justify this argument.

A. The prior

Following [6], we assume that the multi-object posterior before estimation cycle k is given by a p.g.fl. of the form

$$G_{k-1}[h] = G_{k-1}^{\text{ppp}}[h]G_{k-1}^{\text{mbm}}[h] \qquad (9)$$

where

$$G_{k-1}^{\text{ppp}}[h] = \exp\left(v_{k-1}[h-1]\right) \qquad (10)$$

$$G_{k-1}^{\text{mbm}}[h] = \sum_{a \in \mathcal{A}_k} \prod_{i \in \mathcal{T}_{k-1}} w_{k-1}^{i,a^i} G_{k-1}^{i,a^i}[h] \qquad (11)$$

$$G_{k-1}^{i,a^i}[h] = 1 - r_{k-1}^{i,a^i} + r_{k-1}^{i,a^i} f_{k-1}^{i,a^i}[h] \qquad (12)$$

under the constraint that

$$\sum_{a \in \mathcal{A}_k} w_{k-1}^a = \sum_{a \in \mathcal{A}_k} \prod_{i \in \mathcal{T}_{k-1}} w_{k-1}^{i,a^i} = 1. \qquad (13)$$

This constraint ensures that each w_{k-1}^a can be viewed as the probability of the W-hypothesis a. The generality of this form should not be underestimated: The output of any MHT, JPDA, JIPDA, PHD filter or MeMBer filter can be expressed in this manner.

To expand the multi-Bernoulli mixture (11) into a sum over R-hypotheses, we notice that for any W-hypothesis a there may exist several R-hypotheses b. An expansion into R-hypothesis is triggered by every track i for which the existence probability r_{k-1}^{i,a^i} is non-integer. It can be seen from the expressions in [6] that under the no-death assumption, any such track will either be a newborn track, or a track which never has received a second measurement after its root measurement.

To define the hypothesis expansion in mathematical terms, we define the following index sets for a given $a \in \mathcal{A}_{k-1}$:

$$V(a) = \left\{ i \in \mathcal{T}_{k-1} \text{ such that } 0 < r_{k-1}^{i,a^i} < 1 \right\} \qquad (14)$$

$$E(a) = \left\{ i \in \mathcal{T}_{k-1} \text{ such that } r_{k-1}^{i,a^i} = 1 \right\}. \qquad (15)$$

The expansion then goes over all subsets $S \subseteq V(a)$. That is, any such subset S together with the underlying W-hypothesis a constitutes an R-hypothesis b. While there may be several R-hypotheses for each W-hypothesis, there is one and only one W-hypothesis corresponding to each R-hypothesis.

PROPOSITION 1. If an R-hypothesis b is given, then the corresponding W-hypothesis a follows uniquely.

Proof. Any R-hypothesis b generated by the W-hypothesis a arises due to the replacement of zero or more root measurements with zeros. To construct the reverse mapping we need to identify the missing root measurements, so that requirement R1 is satisfied. This can only be done in one particular way, because root measurements only occur in designated places, given by the index function $L(j;k)$. $\qquad\square$

Thus, given an R-hypothesis b, we can denote its underlying W-hypothesis by $a[b]$. Also, let $S[b]$ denote the subset of $V(a)$

that yields b. The multi-Bernoulli mixture can then be written as a sum over R-hypotheses as follows

$$G_{k-1}^{\text{mbm}}[h] = \sum_{b \in \mathcal{B}_{k-1}} w_{k-1}^b \prod_{i \in \mathcal{T}_{k-1}} H_{k-1}^{i,b^i}[h] \qquad (16)$$

where

$$w_{k-1}^b = w_{k-1}^{a[b]} \prod_{i \in V(a) \setminus S[b]} \left(1 - r_{k-1}^{i,a^i[b]}\right) \prod_{i \in S[b]} r_{k-1}^{i,a^i[b]} \qquad (17)$$

$$H_{k-1}^{i,b^i}[h] = \begin{cases} f_{k-1}^{i,a^i[b]}[h] & \text{if } i \in S[b] \bigcup E(a[b]) \\ 1 & \text{otherwise} \end{cases} \qquad (18)$$

and where $\mathcal{B}_{k-1}$ denotes the set of all R-hypotheses generated by expansion of W-hypotheses in $\mathcal{A}_{k-1}$.

REMARK 2. We see that all the track-components in (18) are Bernoulli p.g.fl.'s with unity or zero existence probabilities. Thus, all existence uncertainty has been subsumed by the different R-hypotheses.

REMARK 3. The quantity w_{k-1}^b will automatically satisfy $\sum_{b \in \mathcal{B}_{k-1}} w_{k-1}^b = 1$ insofar as the normalization (13) holds for the original multi-Bernoulli mixture, and gives the probability of b being the correct association. In order to get separate track weights w_{k-1}^{i,b^i} we are free to factorize w_{k-1}^b into n_{k-1} factors in any way we may wish.

B. The prediction

The transition of the prior p.g.fl. $G_{k-1}[h]$ into the predicted p.g.fl. $G_{k|k-1}[h]$ is described in [6] and [12] in terms of W-hypotheses. Under the given assumptions, the transition can also be expressed in terms of R-hypotheses. In particular, zero death probability ensures that the MeMBer-component remains a mixture over the same R-hypotheses. The prediction equations can be found as a special case of Theorem 1 in [6] as

$$v_{k|k-1}(x_k) = \mu_k(x)$$
$$+ \int f_x(x_k|x_{k-1})v_{k-1}(x_{k-1})\mathrm{d}x_{k-1} \qquad (19)$$

$$f_{k|k-1}^{i,b^i}(x_k) = \int f_x(x_k|x_{k-1})f_{k-1}^{i,b^i}(x_k)\mathrm{d}x_{k-1} \qquad (20)$$

$$w_{k|k-1}^{i,b^i} = w_{k-1}^{i,b^i} \qquad (21)$$

$$r_{k|k-1}^{i,b^i} = r_{k-1}^{i,b^i} \in \{0,1\}. \qquad (22)$$

The predicted p.g.fl. is then

$$G_{k|k-1}[h] = G_{k|k-1}^{\text{ppp}}[h]G_{k|k-1}^{\text{mbm}}[h] \qquad (23)$$

where

$$G_{k|k-1}^{\text{ppp}}[h] = \exp\left(v_{k|k-1}[h-1]\right) \qquad (24)$$

$$G_{k|k-1}^{\text{mbm}}[h] = \sum_{b \in \mathcal{B}_{k-1}} \prod_{i \in \mathcal{T}_{k-1}} w_{k|k-1}^{i,b^i} H_{k|k-1}^{i,b^i}[h] \qquad (25)$$

and where the Bernoulli functional $H_{k|k-1}^{i,b^i}[h]$ is given as in (18) with f_{k-1}^{i,b^i} replaced by $f_{k|k-1}^{i,b^i}$.

REMARK 4. The prediction of the Poisson component has no well-defined counterpart in Reid's MHT. The remainder of the prediction is entirely equivalent to the prediction step in Reid's MHT under Gaussian-linear kinematics.

C. The posterior in terms of W-hypotheses

Observing a set of measurements $Z_k = \{z_k^1, \ldots, z_k^{m_k}\}$ causes the set of predicted hypotheses to be expanded into a larger set of posterior hypotheses. Theorem 2 in [6] provides expressions for the elements of this expansion in terms of W-hypotheses, and we repeat this formulation below, adapted to our assumptions and notation.

The posterior p.g.fl. is of the form

$$G_k[h] = G_{k|k-1}^{\text{ppp}}[h] G_k^{\text{mbm}}[h] \tag{26}$$

where

$$G_k^{\text{ppp}}[h] = \exp\left(v_k[h-1]\right) \tag{27}$$

$$G_k^{\text{mbm}}[h] \propto \sum_{a \in \mathcal{A}_k} \prod_{i \in \mathcal{T}_k} w_k^{i,a^i} G_k^{i,a^i}[h] \tag{28}$$

$$G_k^{i,a^i}[h] = 1 - r_k^{i,a^i} + r_k^{i,a^i} f_k^{i,a^i}[h]. \tag{29}$$

In the formulation of W-hypotheses, a new track is established with root in *each* measurement in Z_k. For every track number i, we have four possible cases, governed by the posterior W-hypothesis a. First, a^i may hypothesize that the target in track number i does not exist. Second, a^i may hypothesize that the target in track number i is possibly detected for the first time. We refer to this as a newborn track. In this case, the target may or may not exist. If it does not exist, then z_k^{i,a_k^i} is a false alarm under a. Third, a^i may hypothesize that while the target in track number i previously was detected, it was not detected at the current time step. In this case the target must exist due to the no-death assumption. Fourth, a^i may hypothesize that the target in track number i was both detected previously and at the current time step, in which case it exists.

Using the matrix-based formulation of W-hypotheses proposed in [12] (see Section III), these possibilities can be accounted for by partitioning the track indices $i \in \{1, \ldots, n_k\}$ into four sets:

$$N(a) = \{i \text{ s.t. } a_{1:k}^i = \mathbf{0}_{k \times 1}\}$$
$$B(a) = \{i \text{ s.t. } a_{1:k-1}^i = \mathbf{0}_{k-1 \times 1} \text{ and } a_k^i > 0\}$$
$$M(a) = \{i \text{ s.t. } a_k^i = 0 \text{ and } a_l^i > 0 \text{ for some } l < k\}$$
$$D(a) = \{i \text{ s.t. } a_k^i > 0 \text{ and } a_l^i > 0 \text{ for some } l < k\}. \tag{30}$$

For the nonexistence case $i \in N(a)$ we have

$$w_k^{i,a^i} = 1, \quad r_k^{i,a^i} = 0.$$

For the new target case $i \in B(a)$ we have

$$w_k^{i,a^i} = \lambda + P_{\text{D}} v_{k|k-1}[f_z(z_k^{a_k^i}|\tilde{x}_k)] \tag{31}$$

$$r_k^{i,a^i} = \frac{P_{\text{D}} v_{k|k-1}[f_z(z_k^{a_k^i}|\tilde{x}_k)]}{\lambda + P_{\text{D}} v_{k|k-1}[f_z(z_k^{a_k^i}|\tilde{x}_k)]} \tag{32}$$

$$f_k^{i,a^i}(x_k) = \frac{f_z(z_k^{a_k^i}|x_k) v_{k|k-1}(x_k)}{v_{k|k-1}[f_z(z_k^{a_k^i}|\tilde{x}_k)]}. \tag{33}$$

For the misdetection case $i \in M(a)$ we have

$$w_k^{i,a^i} = w_{k|k-1}^{i,a^i}(1 - P_{\text{D}}) \tag{34}$$

$$r_k^{i,a^i} = 1 \tag{35}$$

$$f_k^{i,a^i}(x_k) = f_{k|k-1}^{i,a^i}(x_k). \tag{36}$$

Notice that the existence probability remains unity in this case under our assumptions. This is in contrast to the more general case treated in [6], where the existence probability may decrease since a misdetection can be taken as evidence for target death. For the measurement update case $i \in D(a)$ we have

$$w_k^{i,a^i} = w_{k|k-1}^{i,a^i} P_{\text{D}} f_{k|k-1}^{i,a^i}[f_z(z_k^{a_k^i}|\tilde{x}_k)] \tag{37}$$

$$r_k^{i,a^i} = 1 \tag{38}$$

$$f_k^{i,a^i}(x_k) = \frac{f_z(z_k^{a_k^i}|x_k) f_{k|k-1}^{i,a^i}(x_k)}{f_{k|k-1}^{i,a^i}[f_z(z_k^{a_k^i}|\tilde{x}_k)]}. \tag{39}$$

The posterior Poisson component, which accounts for undiscovered targets, is given by the intensity function

$$v_k(x) = (1 - P_{\text{D}}) v_{k|k-1}(x). \tag{40}$$

REMARK 5. We see that if there is no hypothesis-conditional existence uncertainty in the predicted multi-Bernoulli mixture, then hypothesis-conditional existence uncertainty in the posterior multi-Bernoulli mixture is only present in tracks with root measurements in the last measurement scan Z_k. These are the tracks belonging to $B(a)$. To translate the posterior to a form similar to Reid's MHT, we therefore need to expand the part of the product in (28) which goes over newborn tracks.

D. The posterior in terms of R-hypotheses

Let $\mathcal{B}_k$ denote the set of R-hypotheses generated by relaxing Requirement R1 for the set $\mathcal{A}_k$ of posterior W-hypotheses, and let $b \in \mathcal{B}_k$. We define the following quantities

$$w_k^b = w_{k|k-1}^{b_{1:k-1}} \lambda^{\varphi} P_{\text{D}}^{\delta}(1 - P_{\text{D}})^{\xi - \delta} \prod_{i \in \mathcal{T}_k} q_k^{i,b^i} \tag{41}$$

$$q_k^{i,b^i} = \begin{cases} 1 & \text{if } i \in N(b) \bigcup M(b) \\ v_{k|k-1}[f_z(z_k^{b_k^i}|\tilde{x}_k)] & \text{if } i \in B(b) \\ f_{k|k-1}^{i,b^i}[f_z(z_k^{b_k^i}|\tilde{x}_k)] & \text{if } i \in D(b). \end{cases} \tag{42}$$

$$\varphi = \text{Number of false alarms according to } b$$
$$\delta = \text{Number of detected targets according to } b$$
$$\xi = \text{Number of existing targets according to } b. \tag{43}$$

The index sets $B(b)$, $N(b)$, $M(b)$ and $D(b)$ are defined as in (30), with a replaced by b. Furthermore, since every R-hypothesis b belongs to a well-defined W-hypothesis $a[b]$, we also define

$$f_k^{i,b^i}[h] = \begin{cases} f_k^{i,a^i[b]}[h] & \text{if } a^i[b] = b^i \\ 1 & \text{otherwise.} \end{cases} \quad (44)$$

PROPOSITION 2. Using the entities defined in (41) - (44), the multi-Bernoulli mixture component of the posterior p.g.fl. can be written as a sum over R-hypotheses as follows

$$G_k^{\mathrm{mbm}}[h] \propto \sum_{b \in \mathcal{B}_k} w_k^b \prod_{i \in \mathcal{T}_k} f_k^{i,b^i}[h]. \quad (45)$$

Proof. Following the observations in Remark 5, we expand the product of Bernoulli p.g.fl.'s over the newborn tracks $i \in B(a)$. Let us first notice that a simple algebraic manipulation yields

$$\begin{aligned} 1 - r_k^{i,a^i} &= 1 - \frac{P_{\mathrm{D}} v_{k|k-1}[f_z(z_k^{a_k^i}|\tilde{x}_k)]}{\lambda + P_{\mathrm{D}} v_{k|k-1}[f_z(z_k^{a_k^i}|\tilde{x}_k)]} \\ &= \frac{\lambda}{\lambda + P_{\mathrm{D}} v_{k|k-1}[f_z(z_k^{a_k^i}|\tilde{x}_k)]} \end{aligned} \quad (46)$$

This allows some cancellations to be made so that the product over newborn tracks becomes

$$\begin{aligned} &\prod_{i \in B(a)} w_k^{i,a^i} G_k^{i,a^i}[h] \\ &= \prod_{i \in B(a)} \left(\lambda + P_{\mathrm{D}} v_{k|k-1}[f_z(z_k^{a_k^i}|\tilde{x}_k)h(\tilde{x}_k)] \right) \\ &= \sum_{S \subseteq B(a)} \lambda^{|B(a) \setminus S|} P_{\mathrm{D}}^{|S|} \prod_{i \in S} v_{k|k-1}[f_z(z_k^{a_k^i}|\tilde{x}_k)h(\tilde{x}_k)]). \end{aligned} \quad (47)$$

With this, the full expression for the multi-Bernoulli mixture becomes

$$\begin{aligned} G_k^{\mathrm{mbm}}[h] &\propto \sum_{a \in \mathcal{A}_k} \sum_{S \subseteq B(a)} \left(\prod_{i \in M(a) \bigcup D(a)} w_k^{i,a^i} f_k^{i,a^i}[h] \right) \\ &\times \left(\lambda^{|B(a) \setminus S|} P_{\mathrm{D}}^{|S|} \prod_{i \in S} f_k^{i,a^i}[h] \right) \end{aligned} \quad (48)$$

What remains in order to arrive at (45) is a small exercise in bookkeeping:

- Counting the numbers of P_{D}-, $(1 - P_{\mathrm{D}})$- and λ-factors in the index subsets $D(a)$, $M(a)$, $B(a)$ and S.
- Recalling that any combination of a and S yields a particular b, and thus replacing a's with b's, in accordance with (44). Also notice that $B(b) = S$, while $N(a) \subseteq N(b)$.

For brevity these details are omitted. $\square$

REMARK 6. The significance of this result is that we obtain a closed form expression, namely (41), for the probabilities of R-hypotheses, derived in the framework of FISST. Even without further approximations or assumptions, the similarity with (16) in [2] should be evident.

The p.g.fl.-representation (45) translates into a multiobject density of the form

$$\sum_{b \in \mathcal{B}_k} w_k^b \sum_{\sigma} \prod_{i \in \mathcal{T}_k} f_k^{i,b^i}(x_k^{\sigma(i)}) \quad (49)$$

where σ ranges over all permutations of $\mathcal{T}_k$. This density contains the available information about the set of discovered targets at time k, and it can be constructed from hypothesis probabilities and hypothesis-conditional pdf's. The summation over σ arises due to the set-derivative (2), and ensures that the multiobject density is permutation-invariant. *For a given association hypothesis b, we can use any of the summands in the sum over σ as a representative of its hypothesis-conditional pdf.* The form of (49) illustrates that target labels have no impact on the tracking results in the standard multitarget Bayes filter. See [18] for more extensive discussion of this observation. Even though it therefore appears redundant to put labels on targets in the *standard* multitarget Bayes filter, it may still be useful to put labels on tracks, i.e., on columns of a or b. Also, target labels may have a meaningful role to play in *non-standard* tracking applications, such as fusion between radar and transponder signals in air traffic control.

VII. GAUSSIAN-LINEAR KINEMATICS AND REID'S MHT

A. Gaussian-linear kinematic assumptions

For the special case of a Gaussian-linear tracking problem we make the following standard assumptions

$$f_x(x_k|x_{k-1}) = \mathcal{N}(x_k; Fx_{k-1}, Q) \quad (50)$$
$$f_z(z_k^j|x_k^i) = \mathcal{N}(z_k^j; Hx_k^i, R) \text{ if } z_k^j \text{ is gen. by } x_k^i \quad (51)$$

in addition to the assumptions already listed in Section V. Furthermore, we assume that the state vector and observation matrix are as in Section VII of [2]

$$\begin{aligned} x_k &= [\rho_{x,k}, \rho_{y,k}, v_{x,k}, v_{y,k}]^{\mathsf{T}} \\ H &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \end{aligned} \quad (52)$$

where ρ denotes position and v denotes velocity.

B. The birth model

When Gaussian-linear kinematics are used in the general machinery so far presented, the only difference from Reid's MHT is with regard to the treatment of newborn targets. The construction of the birth process was only sketched in colloquial terms in [2 p. 848], and summarized by a single Poisson intensity denoted β_{NT}. Comparing this with the standard model used here, it is not clear whether β_{NT} should be taken as referring to the Poisson intensity of the unknown target density $v(x)$ or the bona-fide birth density $\mu(x)$. In this paper, we choose the former option since this obviates the need for dealing with both quantities in a framework that only allows for one such quantity. Consequently, we do not discuss estimation of $v(x)$ any further in the sequel. Instead we assume

that it is given in a form that resembles the description and recommendations in [2] as much as possible:

$$v(x) = \beta_{\text{NT}} \chi_S(Hx) \mathcal{N}(H^* x \,;\, \mathbf{0}, P_v).$$

Here $\chi_S(\cdot)$ is the indicator function of the surveillance region, P_v is a diagonal 2×2 matrix representing *a priori* knowledge of target velocity, and

$$H^* = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{53}$$

such that $H^* x_k = [v_{x,k}, v_{y,k}]^{\mathsf{T}}$.

C. Comparison with Reid's MHT

Under these assumptions, we get the following expressions for the functionals involved in the hypothesis score w_k^b:

$$\begin{aligned} v_{k|k-1}[f_z(z_k^{b_k^i}|\tilde{x})] &= \beta_{\text{NT}} \int \mathcal{N}(z_k^{b_k^i} \,;\, Hx, R) \\ &\qquad \mathcal{N}(H^* x \,;\, \mathbf{0}, P_v) \, \mathrm{d}x \\ &= \beta_{\text{NT}} \end{aligned} \tag{54}$$

$$\begin{aligned} f_{k|k-1}^{i,b^i}[f_z(z_k^{b_k^i}|\tilde{x}_k)] &= \int \mathcal{N}(z_k^{b_k^i} \,;\, Hx, R) \\ &\qquad \mathcal{N}(x \,;\, x_{k|k-1}^{i,b^i}, P_{k|k-1}^{i,b^i}) \, \mathrm{d}x \\ &= \mathcal{N}(z_k^{b_k^i} \,;\, Hx_{k|k-1}^{i,b^i}, S_{k|k-1}^{i,b^i}) \end{aligned} \tag{55}$$

where $S_{k|k-1}^{i,b^i} = H P_{k|k-1}^{i,b^i} H^{\mathsf{T}} + R$, and where $x_{k|k-1}^{i,b^i}$ and $P_{k|k-1}^{i,b^i}$ are sufficient statistics for the pdf $f_{k|k-1}^{i,b^i}(x)$. Inserting these expressions into (41) yields

$$\begin{aligned} w_k^b &= w_{k|k-1}^{b_{1:k-1}} \lambda^\varphi P_{\text{D}}^\delta (1 - P_{\text{D}})^{\xi - \delta} \beta_{\text{NT}}^{|B(b)|} \\ &\quad \times \prod_{i \in D(b)} \mathcal{N}(z_k^{b_k^i} \,;\, Hx_{k|k-1}^{i,b^i}, S_{k|k-1}^{i,b^i}). \end{aligned} \tag{56}$$

This expression is identical to Reid's recursive hypothesis probability formula (16) in [2].

VIII. Conclusion

We have shown that Reid's MHT can be established from the framework of FISST. In particular, we have derived Reid's recursive hypothesis probability formula by means of a representation in terms of multi-Bernoulli mixtures, with basis in [6].

The only discrepancies between the classical MHT approach and the modern RFS approach lie within the treatment of newborn and unobserved targets. Our derivation of Reid's formula requires the distribution of unknown targets to be Poisson and spatially constant, and known by means of some procedure outside of the MHT framework.

Concerning future and ongoing research, it is possible to conduct a similar analysis purely in terms of multiobject densities. Also, the analysis can be done under the interpretation of β_{NT} as a standard birth model, instead of the unknown target density. This yields a formulation of the multitarget Bayes filter which deviates slightly from Reid's MHT.

It remains debatable whether FISST or MHT gives the most general formulation. On the one hand we did indeed develop Reid's MHT within FISST. On the other hand the association hypotheses contain more information than what is directly encoded in the multiobject density (49).

References

[1] C. L. Morefield, "Application of 0-1 integer programming to multitarget tracking problems," *IEEE Transactions on Automatic Control*, vol. 22, no. 3, pp. 300–10, Jan. 1977.

[2] D. Reid, "An algorithm for tracking multiple targets," *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 843–854, Dec. 1979.

[3] R. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Norwood, MA, USA: Artech House, 2007.

[4] ——, "Multitarget Bayes filtering via first-order multitarget moments," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1152–1178, May 2003.

[5] ——, "PHD filters of higher order in target number," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 4, pp. 1523–1543, October 2007.

[6] J. Williams, "Marginal multi-Bernoulli filters: RFS derivation of MHT, JIPDA, and association-based MeMBer," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 3, pp. 1664–1687, July 2015.

[7] S. Challa, M. R. Morelande, D. Musicki, and R. J. Evans, *Fundamentals of object tracking*. Cambridge University Press, 2011.

[8] M. Ulmke, O. Erdinc, and P. K. Willett, "Gaussian mixture cardinalized PHD filter for ground moving target tracking," in *Proceedings of Fusion 2007*, Quebec, Canada, July 2007.

[9] S. Mori, C.-Y. Chong, and K. C. Chang, "Three formalisms of multiple hypothesis tracking," in *Proceedings of FUSION*, Heidelberg, Germany, Jul. 2016.

[10] R. Streit, "How I learned to stop worrying about a thousand and one filters and love analytic combinatorics," in *Proceedings of IEEE Aerospace Conference*, Big Sky, MT, USA, Mar. 2017.

[11] E. Brekke and M. Chitre, "A multi-hypothesis solution to data association for the two-frame SLAM problem," *The International Journal of Robotics Research*, vol. 34, no. 1, pp. 43–63, Jan. 2015.

[12] E. Brekke, B. Kalyan, and M. Chitre, "A novel formulation of the Bayes recursion for single-cluster filtering," in *Proceedings of IEEE Aerospace Conference*, Big Sky, MT, USA, Mar. 2014.

[13] A. J. Swain, "Group and extended target tracking with the probability hypothesis density filter," Ph.D. dissertation, Heriot Watt University, Edinburgh, United Kingdom, Nov. 2013.

[14] T. Kurien, "Issues in the design of practical multitarget tracking algorithms," in *Multitarget-Multisensor Tracking: Applications and Advances*, Y. Bar-Shalom and W. D. Blair, Eds. Artech House, 1990, ch. 3, pp. 219–245.

[15] S. Mori, C.-Y. Chong, E. Tse, and R. Wishner, "Tracking and classifying multiple targets without a priori identification," *IEEE Transactions on Automatic Control*, vol. 31, no. 5, pp. 401–409, May 1986.

[16] B.-T. Vo and B.-N. Vo, "Labeled random finite sets and multi-object conjugate priors," *IEEE Transactions on Signal Processing*, vol. 61, no. 13, pp. 3460–3475, Jul. 2013.

[17] Y. Bar-Shalom, P. K. Willett, and X. Tian, *Tracking and Data Fusion: A Handbook of Algorithms*. Storrs, CT, USA: YBS Publishing, 2011.

[18] E. H. Aoki, P. K. Mandal, L. Svensson, Y. Boers, and A. Bagchi, "Labeling uncertainty in multitarget tracking," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 3, pp. 1006–1020, June 2016.